PREDICTING MUSIC EMOTION WITH SOCIAL MEDIA DISCOURSE

From Reddit, Twitter, and YouTube conversations, we predict emotive responses to music.

MACHINE UNDERSTANDING OF MUSIC EMOTION

Recently, music emotion recognition has been of interest in the field of music information retrieval. Current methods generally rely on human annotators to rate samples of music through crowdsourced survey platforms. These surveys can be expensive and difficult to glean useful information from due to the subjective nature of emotion annotation. Our hypothesis is that the conversations people have online about specific songs contain semantic information which a machine learning model could be trained on to predict the emotion elicited in a listener by a song.

Figure 1: A wordcloud describing the top 200 emotive words in our dataset.

Dataset	Songs	Posts	Words
AMG1608[1]	1608	351,562	4,725,419
DEAM[4]	1803	67,764	832,886
PmEmo[5]	768	227,054	2,886,369

Table 1: A summary of the datasets used and the social media data aggregated from them.

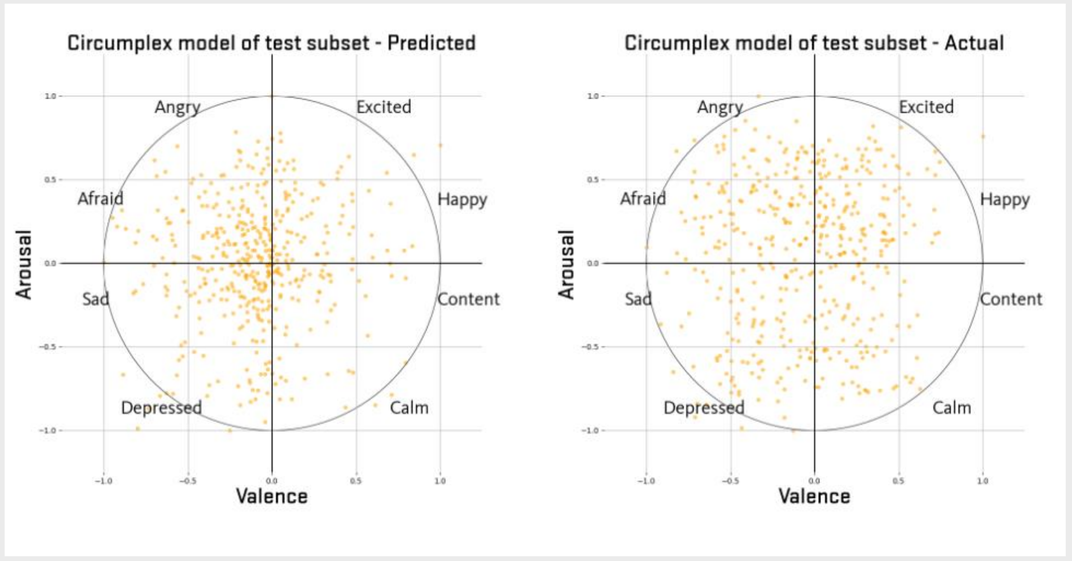

Figure 2: AMG1608 test subset valence-arousal plot, comparing true labels to predicted labels.

METHODS

We analyze three music emotion datasets with human-annotated valence and arousal labels from crowdsourced surveys. For each song in a given dataset, we scrape the top 10 submissions related to that track on Reddit, YouTube, and Twitter.

Experiment 1 analyzes individual words in a post for semantic information. We extract affective terms from a given comment using five word-emotive dictionaries[2]. From these emotive unigrams, we calculate the average and standard deviation of mood categories, sentiment, and valence/arousal.

DistilBERT[3], a pre-trained transformer model, allows us to learn with raw comments as input. Experiment 2's deep learning approach enables the recognition of not just individual affective terms, but the context in which they are used as well.

	Experiment 1		Experiment 2	
Dataset	Valence	Arousal	Valence	Arousal
AMG1608	0.094	0.295	0.454	0.655
DEAM	0.170	0.198	0.197	0.254
PmEmo	0.597	0.464	0.684	0.583

Table 2: Pearson's correlation coefficients of predicted valence and arousal values.

RESULTS

We train a random forest on our word-level features in Experiment 1. Overall performance was negligible, with the notable exception of songs from the PmEmo dataset, where our predictions exhibited much stronger correlation to ground truth. We believe this to be due to the songs from PmEmo being selected from Billboard Top 100, versus the more obscure tracks used in other music emotive datasets, resulting in extended online discourse about these tracks. This is evidenced by PmEmo's greater ratio of posts to songs.

Experiment 2 averages a 55% improvement over our unigram analysis, with up to a 185% improvement in the case of AMG1608. We observe a less impressive 22% improvement within the DEAM dataset, likely due to a lack of significant social media conversation surrounding these tracks.

DISCUSSION

Our transformer model averages Pearson's correlations between 0.45 and 0.65 in datasets with significant commentary. This indicates the presence of affective semantics in a comment's structure. Though this correlation is weak, it demonstrates the feasibility of predicting music emotion directly from social media discussion. To our knowledge, this is the first experiment to apply social media sentiment analysis in an attempt to estimate music emotion.

ACKNOWLEDGEMENTS

Aidan Beery – Project Lead
Contact: beerya@oregonstate.edu

Dr. Patrick Donnelly - Advisor
Contact: donnellp@oregonstate.edu
Website: https://soundbendor.org/

REFERENCES

[1] Y.-A. Chen *et. al.*, "The AMG1608 dataset for music emotion recognition," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, Apr. 2015, pp. 693–697.

[2] S. Mohammad, "Word Affect Intensities," *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference*, 2018.

[3] V. Sanh *et. al.*, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," in 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing - NeurIPS 2019, Oct. 2019.

[4] Aljanaki A, Yang Y-H, Soleymani M (2017) Developing a benchmark for emotional analysis of music. PLoS ONE 12(3): e0173392. https://doi.org/10.1371/journal.pone.0173392

[5] K. Zhang *et. al.*, "The PMEmo Dataset for Music Emotion Recognition," in Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, Jun. 2018, pp. 135–142.

Oregon State University

soundbendor lab